DOCUMENT RESUME

ED 397 534 EA 027 930

AUTHOR Khattri, Nidhi; And Others
TITLE Assessment of Student Performance. Volume III: Technical Appendix. Studies of Education Reform.
INSTITUTION Pelavin Research Inst., Washington, DC.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 1 Nov 95
CONTRACT RR91172004
NOTE 23p.; For Volumes I-III of this particular study, see EA 027 928-930. For all 12 final reports (36 volumes) in this series of studies, see EA 027 926-961.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Accountability; Case Studies; Data Analysis; Data Collection; *Educational Assessment; *Evaluation Criteria; Instructional Improvement; Longitudinal Studies; *Organizational Objectives; *Performance; Secondary Education; *Student Evaluation; Teaching Methods
IDENTIFIERS *Studies of Education Reform (OERI)

ABSTRACT

The emphasis on using performance assessments to support systematic state-, district-, or school-wide purposes is a recent development in the assessment-reform movement. A 3-year case study sought to describe the status of assessment reform in United States education systems and offer recommendations for policy and future research. The research design, which used a qualitative, case-study approach to collect data about performance assessments and their impacts at the school level is described. Researchers visited 16 sites in spring 1994 and revisited 7 sites in spring 1995. The report describes the analysis of background literature (a review of theoretical, empirical-research, and policy papers) and the case study methodology, which examined state-, district-, school-, and national-level assessments. Data for the case studies were gathered through document analysis; telephone interviews; classroom observation; and site interviews with state Department of Education representatives, school district representatives, principals and assistant principals, teachers, other school staff, students, parents, and school board members. The report also describes the criteria for sample selection, a description of the sample, and data-analysis procedures (within-case and cross-case). Seven exhibits and a list of the research questions are included. (LMI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 397 534

STUDIES OF EDUCATION REFORM: ASSESSMENT OF STUDENT PERFORMANCE

FINAL REPORT

VOLUME III: TECHNICAL APPENDIX

Project Director: Michael B. Kane

Authored By:

Nidhi Khattri
Alison L. Reeve
Michael B. Kane

November 1, 1995

Submitted To:

Office of Educational Research and Improvement
U.S. Department of Education

Submitted By:

Pelavin Research Institute
1000 Thomas Jefferson Street, Suite 400
Washington, DC 20007

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

EA 027 930

BEST COPY AVAILABLE

STUDIES OF EDUCATION REFORM: ASSESSMENT OF STUDENT PERFORMANCE

FINAL REPORT

VOLUME III: TECHNICAL APPENDIX

Project Director: Michael B. Kane

Authored By:

Nidhi Khattri
Alison L. Reeve
Michael B. Kane

November 1, 1995

Submitted To:

Office of Educational Research and Improvement
U.S. Department of Education

Submitted By:

Pelavin Research Institute
1000 Thomas Jefferson Street, Suite 400
Washington, DC 20007

PREFACE

Studies of Education Reform: Assessment of Student Performance is the result of a research project conducted by Pelavin Research Institute (PRI), an affiliate of the American Institutes for Research (AIR), under a contract with the Office of Educational Research and Improvement (OERI) of the U. S. Department of Education (Contract Number RR91172004). In 1991, OERI issued a request for proposals entitled "Studies of Education Reform." Twelve studies were incorporated under this general heading, each reflecting some aspect of the reform movement that had placed education at the forefront of the national agenda in the late 1980s. PRI was awarded a three-year contract to study assessment reform, which we interpreted to mean the contribution of performance-based, non-multiple choice assessments to education reform.

Contractors for all 12 reform studies were required to hold a national conference within the initial year of their study and to commission papers on important aspects of the reform topic. PRI, in collaboration with the OERI study of curriculum reform (conducted by Ronald Anderson of the University of Colorado), held a national conference on performance assessment and curriculum reform as a pre-session to the Annual Student Assessment Conference, organized by the Education Commission of the States, in Boulder, Colorado, in June of 1992. The assessment component of the pre-session conference included discussions of the content of nine commissioned papers that are to appear in the book, *Implementing Performance Assessments: Promises, Problems, and Challenges* (Kane & Mitchell, in press).[1] The papers, the conference attendees' insights, and OERI's research questions helped us refine our study's intellectual and methodological framework.

The larger and more significant context for this study was the increasing commitment across the nation to performance assessment as a reform strategy. For example, California spearheaded the reform movement with statewide open-ended mathematics assessments in the late 1980s, and Vermont followed suit with its first, statewide portfolio assessments. In the late 1980s and early 1990s, other states, districts, and schools also began developing and implementing performance-based assessments.

In the subsequent two years, we visited 16 schools across the country (of which we revisited seven) that were participating in the development or implementation of performance assessments as a result of national, state, district, or local assessment reform initiatives. In addition, we presented papers based upon our initial study findings at the American Educational Research Association conference in April, 1995, and the Council of Chief State School Officers' conference on Large-Scale Assessment in June, 1994 and 1995.

This volume, the third in a three-volume series, identifies the specific study objectives and provides an overview of the study design and objectives. It discusses our research design, sample selection criteria, data collection strategies, and data analysis. The results of the project are contained in Volume I: Findings and Conclusions and case studies of the schools in our sample are contained in Volume II: Case Studies.

---

[1] All royalties resulting from the sales of this book will be contributed to the Leigh Burstein Memorial Fund, administered by the University of California at Los Angeles Foundation.

ACKNOWLEDGMENTS

The support and participation of several individuals and institutions made this study possible. We are grateful to the state and district education agency personnel and to the principals, teachers, students, parents, and school board members who devoted large amounts of their precious time to answer our questions and to guide us in understanding the issues related to assessment reform. We regret that our promise of anonymity leaves them without the public recognition they deserve.

We also are grateful to David Sweet, the OERI monitor of *Studies of Education Reform: Assessment of Student Performance*, whose incisive feedback guided our work at every step, and to Ruth Mitchell, who served as Project Manager during the study's initial stages and as Senior Advisor during the later years. Her steady interest in and support of the project was invaluable.

This study also benefitted from the participation and general support of the members of our Advisory Committee, including: Dr. George Elford, Educational Testing Service (retired); Dr. Lauren Resnick, Learning Research and Development Center, University of Pittsburgh; Dr. Ramsey Selden, Council of Chief State School Officers (now of the Education Statistical Services Institute); Dr. Karen Sheingold, Educational Testing Service; Dr. Loretta Shepard, Department of Education, University of Colorado; and Dr. Grant Wiggins, Center on Learning, Assessment, and School Structure. In 1991-92, the Advisory Committee members also included: Dr. Robert Mislevy, Educational Testing Service, and Mr. Sol Pelavin, American Institutes for Research.

We also wish to acknowledge the capable support and contributions of the following individuals: Rebecca Adamson, Kerry Traylor, Doug Levin, Ray Varisco, Gwen Pegram, Kimberly Cook, Michael Garet, Kimberly Gordon, Rebecca Shulman, and Shelley Kirkpatrick, all at Pelavin Research Institute; and Amy Stempel and Stepahnie Soper, both at the Council for Basic Education.

TABLE OF CONTENTS

PREFACE i

ACKNOWLEDGMENTS iii

TABLE OF CONTENTS v

LIST OF EXHIBITS vii

TECHNICAL APPENDIX:
RESEARCH DESIGN AND METHODOLOGY 1

Objectives 1
Research Design 1
Collection and Analysis of Background Literature 3
Qualitative Research Methodology: A Case Study Approach 3
Sample Selection Criteria 4
Sample Description 5
Data Collection Activities and Instruments 8
Data Analysis 10
Within-Case Analysis 10
Cross-Case Analysis 11
Data Interpretation 12

ATTACHMENT A A-1

LIST OF EXHIBITS

EXHIBIT 1: Conceptual Scheme of Major Objectives 2

EXHIBIT 2: Site Visit Design 4

EXHIBIT 3: 16 Round One Sites 6

EXHIBIT 4: Sample Characteristics 7

EXHIBIT 5: Seven Round Two Sites 8

EXHIBIT 6: Roles and Numbers of Interviewees by Round of Site Visit 9

EXHIBIT 7: First and Second Round Site Visit Observations 10

TECHNICAL APPENDIX:
RESEARCH DESIGN AND METHODOLOGY

Objectives

The major objectives of the three-year longitudinal study ***Studies of Education Reform: Assessment of Student Performance*** are as follows:[1]

Objective 1: *Document and analyze key characteristics of performance assessments;*

Objective 2: *Document and analyze facilitators and barriers in assessment reform; and*

Objective 3: *Document and assess impacts of performance assessments on teaching and learning.*

Our ultimate purpose in this study was to elucidate the status of assessment reform in U.S. education systems and to offer recommendations for policy and future research.

Our approach to meeting the three objectives outlined above was driven by our conceptualization of the relationships among the factors driving and affecting assessment reform and its outcomes. We conceptualized the key characteristics of performance assessments and the facilitators and barriers in assessment reform as interdependent variables that influence teaching and learning processes and student achievement. Exhibit 1 illustrates our conceptual scheme.

Research Design

Our research design employed a qualitative, case-study approach to collecting data about performance assessments and their impacts at the school level. In addition, during the course of the study, we also collected a library of policy, research, historical, and other documents on performance assessments, assessment reform, and education reform in general. Below, we describe the following aspects of our research design:

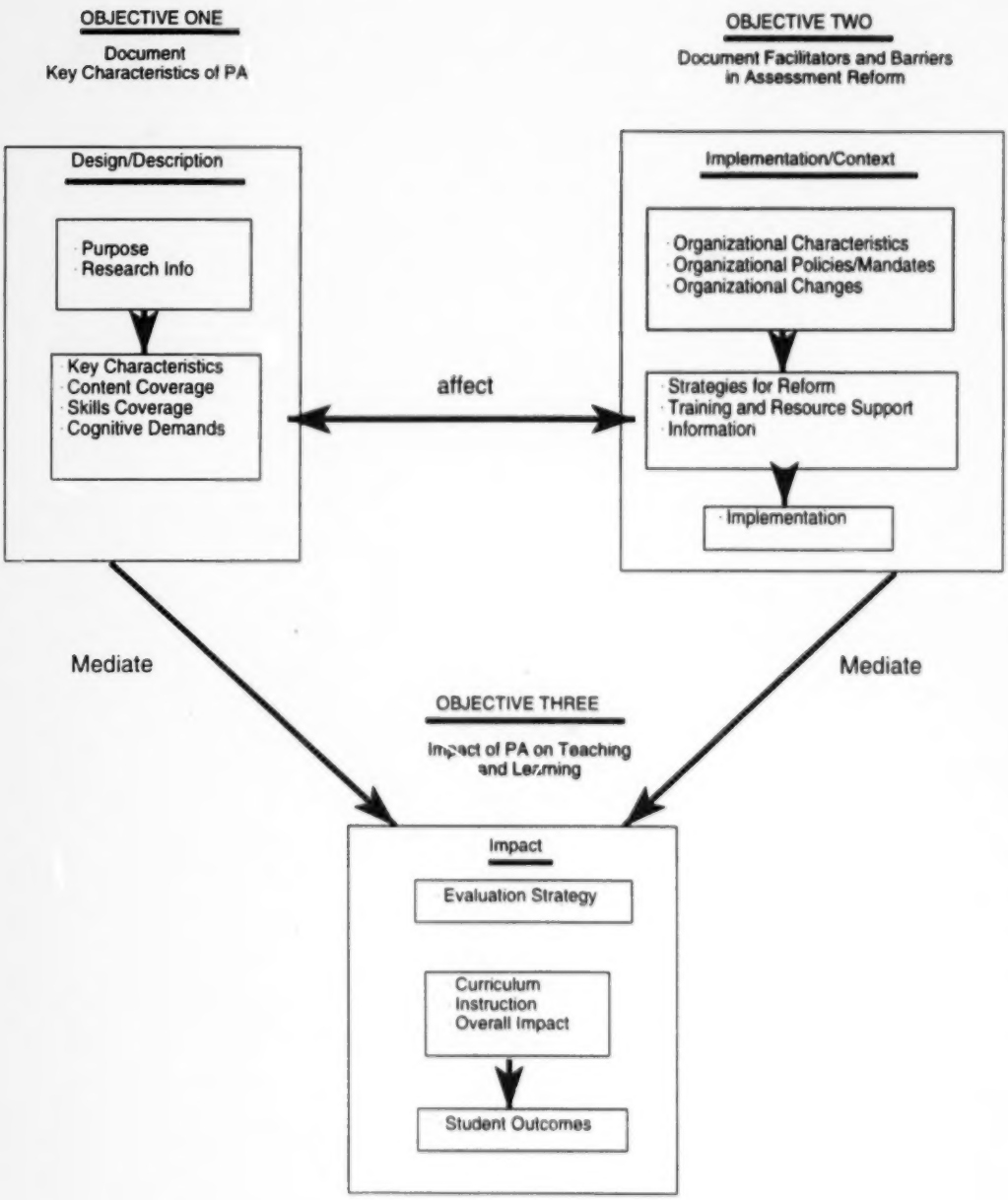
- collection and analysis of background literature,
- the qualitative, case-study methodology,
- sample selection criteria,
- sample description,

---

[1] The specific research questions are presented in Attachment A.

EXHIBIT 1

Conceptual Scheme of Major Objectives



BEST COPY AVAILABLE

- data collection activities, and
- data analysis procedures.

We conclude by pointing out the strengths and weaknesses of the research design employed with respect to meeting the purposes of the study.

Collection and Analysis of Background Literature

We collected the following types of documents continuously over the entire span of our study:

- Theoretical papers on issues dealing with all aspects, including the development and implementation, of performance assessments.
- Empirical research papers on issues regarding effectiveness, equity, psychometrics of performance assessments, and data on student outcomes.
- Policy papers on the development, implementation, and uses of performance assessments.

We culled relevant information from these documents in a literature review at the beginning of the project, in the Fall of 1992, and again in the Spring of 1995, towards the close of the project. We also used these papers to deepen our understanding of the issues related to the development, implementation, and impact of performance assessments and to inform our analyses of the case-study data.

Qualitative Research Methodology: A Case Study Approach

We employed a qualitative, case-study methodology to investigate the development and implementation of performance assessments and their impacts at the school level. We designed a modified time-series approach for gathering data, which enabled us to obtain both cross-sectional and longitudinal data. Cross-sectional data allowed us to make comparative remarks about assessments and school contexts. The longitudinal data allowed us to document the effects of and changes in performance assessments over time within sites.

We selected 16 sites (the definition of "site" for this study encompasses both a performance assessment and a single school at which it is being used) which a team of two researchers visited a single time during a two-day site visit. We then selected a subset of 7 sites, which the team returned to for a second visit (therefore, longitudinal data were collected for only 7 of the 16 sites).

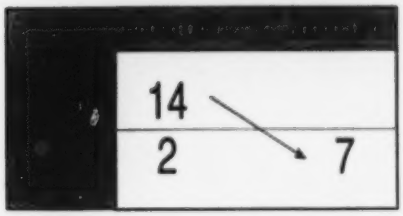
We conducted the first set of site-visits in the Spring of 1994, and the second set of site-visits in the Spring of 1995. (Two of the single time site-visits were conducted in the Spring of 1995). Exhibit 2 shows our site-visit design.

Sample Selection Criteria

As described above, our research design called for two waves of data collection: first round of visits to all sites included in the sample, followed by a second round of visits to a subset of those sites. Below, we describe the criteria we applied to select sites for inclusion in rounds one and two.

EXHIBIT 2

Site Visit Design



Selection Criteria: Round One Sites

The overarching objective of our site selection process was to identify, insofar as possible, a set of school sites that exhibited the range of experiences American schools are having with the development and implementation of performance assessments. For the purposes of our study, we defined a case study site as a single school where a performance assessment was being implemented. The focus of our research was on the assessment and its implementation in the local context. To select the sites, we delineated two sets of criteria – those pertaining to performance assessments and those pertaining to schools.

Performance assessments are marked by a number of variable characteristics, and we attempted to obtain variation in our sample within each characteristic. Selection criteria pertaining to performance assessment characteristics included:

- **Type of Assessment.** Performance assessments come in a variety of forms, including portfolios, on-demand assessments, demonstrations and presentations, and extended projects. We wanted our sample to include assessments that reflected this range so that we might discern variation in effects of assessment type on teaching and learning at the local level.

- **Locus of Development.** The movement toward the use of performance-based assessments is taking place at all levels of educational authority. States, districts, and schools alike are developing and implementing performance assessments. Furthermore, some national-level efforts, such as the New Standards Project and the Coalition of Essential Schools, are also influencing the turn toward performance assessments. Because the purposes, design, and impact of assessments developed at different levels of authority could potentially vary significantly, we wanted our sample to reflect this diversity in locus of development.

- **Status of Implementation.** The entire performance assessment movement is still relatively young. However, performance assessments do vary with respect to their stage in development and implementation. Therefore, we wanted our sample to include assessments with varying status of implementation: developmental and pilot, full-scale implementation, and maintenance.

- **Content Area.** Performance assessments can be used in all subject areas, but the assessments can look different for different subject areas. We wanted our sample to include assessments that focused on a range of subject areas including language arts, mathematics, science, and social science.

Although we were less concerned with school background characteristics (e.g., size, racial and ethnic composition, and socio-economic background of the student body) than we were with the assessment characteristics delineated above, we attempted also to obtain variation across two school characteristics:

- **School Level.** We wanted the sample to include elementary schools, middle schools, and high schools, since performance assessments might affect teachers and students differently at the various levels of schooling.

- **Geographical Diversity.** Because American children are educated in schools in 50 states and the District of Columbia, and because these schools are located in urban, suburban, and rural areas, we wanted the sample to include school sites located in various regions of the country and in communities of varying urbanicity.

As the sample for this qualitative study was to include 16 schools, it is clear that not all combinations of the six factors described above could be included (if, indeed, they even exist). Rather, we aimed to create a sample in which the range of characteristics each of the six criteria was represented.

Selection Criteria: Round Two Sites

We chose a subset of seven sites for a second round of data collection. We based selection of the seven round two sites upon one or more of the following criteria:

- It was anticipated that changes in the performance assessment design or implementation would take place between 1993-94 and 1994-95.

- Our understanding of the effects of assessment reform at the site was less than clear based on one round of data collection and was likely to improve with a second visit.

Sample Description

Sixteen performance assessments at 16 school sites were selected to comprise the study sample. The 16 sites are identified in Exhibit 3.

EXHIBIT 3

16 Round One Sites*

SITES
State-level assessments:
• Arizona Assessment Reform, Manzanita High School
• Kentucky Assessment Reform, Breckenridge Middle School
• Maryland Assessment Reform, Walters Middle School
• New York Regents Portfolios, Hudson High School
• Oregon Assessment Reform, Crandall High School
• Vermont Portfolio Assessments, Maple Leaf Middle School
District-level assessments:
• Harrison School District 2's Performance-Based Curriculum, McGary Elementary School (CO)
• South Brunswick Unified School District's 6th Grade Research Performance Assessment, Windermere Elementary School (NJ)
• Prince William County Public Schools' Applications Assessments, Westgate Middle School (VA)
School-level assessments:
• Language Arts and Math Portfolios, Niños Bonitos Elementary School (CA)
• Primary Learning Record, Park Elementary School (NY)
• Rite of Passage Experience, Thoreau High School (WI)
National-level assessment projects:
• New Standards Project, Ann Chester Elementary School (TX)
• New Standards Project, Noakes Elementary School (IA)
• Coalition of Essential Schools, Cooper Middle School (NM)
• College Board's Pacesetter Mathematics Program, Sommerville High School (MD)

* All schools have been assigned pseudonyms.

Exhibit 4 provides information about the characteristics of each of the 16 sites included in the study sample. As illustrated, these characteristics demonstrate the variation we achieved in our sample with respect to our selection criteria.

EXHIBIT 4

Sample Characteristics

Assessment Criterion	Sample
Type of Assessment	A majority (10 of the 16) of the assessments involved multiple types of assessment tasks. The number of sites using each type of assessment is as follows: 8 on-demand tasks 8 extended performance tasks or projects 10 portfolios 3 demonstrations or presentations 2 teacher observations
Locus of Development	4 national-level assessments (3 national organizations, 4 sites). 6 state-level assessments 3 district-level assessments 3 school-level assessments
Status of Implementation	3 pilot and development 12 full-scale implementation 1 implementation by individual teachers
Content Area	Most (13 of the 16) of the assessments incorporated multiple subject areas. Some were integrated assessments, while others incorporated distinct assessment tasks for various subject areas. Most assessments that focused on multiple subject areas included language arts and mathematics, and some included science and social science. The number of performance assessments focusing on different subject areas is as follows: 11 language arts 8 mathematics 5 science 5 social studies 1 practical living/vocational
School Criterion	**Sample**
Grade Level	6 elementary schools 5 middle schools 5 high schools
Geographical Diversity	As is demonstrated in Exhibit 3, our sample included schools from most regions of the country. One exception was an under representation of southern states: sites in Kentucky and Virginia were included, but no states from the deep south were represented in the sample. This was in part because at the time of sample selection, few southern states were introducing state-level performance assessments. The sample is composed primarily of schools located in urban and suburban areas and small cities. Rural sites are under-represented.

Exhibit 5 identifies the seven sites we selected for a second visit during round two of our data collection.

EXHIBIT 5

Seven Round Two Sites

SITES
State-level assessments:
• Kentucky Assessment Reform, Breckenridge Middle School
• Oregon Assessment Reform, Crandall High School
• Vermont Portfolio Assessments, Maple Leaf Middle School
District-level assessments:
• Prince William County Public Schools' Applications Assessments, Westgate Middle School (VA)
School-level assessments:
• Language Arts and Math Portfolios, Niños Bonitos Elementary School (CA)
• Primary Learning Record, Park Elementary School (NY)
National-level assessment projects:
• Coalition of Essential Schools, Cooper Middle School (NM)

Data Collection Activities and Instruments

Because we were interested in obtaining information about the performance assessment, the educational context within which it was developed and implemented, and the assessment's effects at the local level, we collected documentary, phone interview, and site-visit data.

Documentary Data

Prior to and during each site visit we collected background documentary data about the subject assessment. The available data varied across assessments. Types of data collected include:

- Descriptions of performance assessments,
- Samples of performance assessments,
- Policy documents about the assessments,
- Policy documents about related education reform efforts,
- Evaluation and research reports regarding the assessments, and
- Newspaper reports about the assessments.

These data were collected from state and local education officials, school staff members, and representatives of external groups involved in assessment reform (e.g., the New Standards Project). These data were collected throughout the life of the project.

We also collected documentary data about the school sites we visited. These data included reports describing each school's demographic composition, staff description, financial resources, and other relevant documents.

Phone Interview Data

Prior to each site visit, we also conducted initial telephone interviews with cognizant individuals in state and local education offices, the school site, and external assessment reform organizations. We used an interview protocol tailored to the role of the interviewee and to the performance assessment system under investigation.

Site Visit Data

In the Spring of 1994 we visited the first 14 schools in our sample. In the Spring of 1995 we revisited seven schools and added two new ones to our sample. In total, we conducted 23 site-visits.

Each site visit lasted one-and-a-half to two days and was conducted by a team of two researchers. The researchers interviewed a number of individuals, observed classrooms, and, whenever possible, observed professional development sessions devoted to the development and use of performance assessments, administration of performance assessments, other activities related to the implementation of performance assessments.

Exhibit 6 illustrates the roles and numbers of the individuals we interviewed during our first and second round site visits.

EXHIBIT 6

Roles and Numbers of Interviewees by Round of Site Visit

Role of Interviewee	Number of Individuals Interviewed		
	Round One (16 Sites)	Round Two (7 Sites)	Total
Representatives of the State Department of Education	6	1	7
Representatives of the School District	16	7	23
Principals and Assistant Principals	18	6	24
Teachers	123	40	163
Other School Staff (e.g., librarians, counselors)	3	2	5
Students	76	26	102
Parents	26	11	37
School Board Members	10	4	14

We used semi-structured interview protocols during our site visits. The protocols for both waves of data-collection were quite similar in structure, but wave two protocols contained more probing questions about the use and effects of performance assessments on teaching and learning. All interview protocols appear in the technical appendix to this report.

Researchers also observed various performance assessment-related activities. Exhibit 7 illustrates the types and numbers of assessment-related activities observed.

EXHIBIT 7

First and Second Round Site Visit Observations

Type of Observation	Number
Classroom Instruction - observations ranged in length from about half an hour to a full school day	22
Assessment Administration - including student presentations, group activities preliminary to assessment administration, and administration of on-demand assessment tasks	9
Professional Development Sessions	9
Other - including a team teaching meeting, one school's end of the year "Assessment Night" when students share their portfolios with their parents, training for outside assessors, and a meeting of a District Assessment Committee	4

Data Analysis

The analyses of our data progressed in two overlapping phases. The two phases, within-case data analysis and cross-case data analysis, are detailed below. We utilized *Qualitative Data Analysis* (Miles & Huberman, 1995) as a sourcebook to inform our data analysis methods.

Within-Case Analysis

The first phase of our data analysis consisted of writing case studies of our sites. To reiterate, the definition of our "site" is a performance assessment and its implementation and impact at one particular school. Data from all sources — documents, interviews, and observations — were synthesized in the case-study report. One member of each site-visit team wrote the case study, and the second member reviewed it for accuracy. Next the case-study was sent to the appropriate officials and school personnel for review and comments. Based upon their feedback, we revised the case study write-ups.

Each case study write-up is divided into the following four sections:[2]

- **Section One: School Profile and Introduction.** This section briefly profiles the organizational characteristics of the school, such as demographic data about its students and the community it serves. The section also presents information on the number and roles of the individuals interviewed and the types of observational data collected.

- **Section Two: Description of Performance Assessment.** This section presents a brief history of the development and implementation of the performance assessment. It also presents the key characteristics of the assessment, including purposes, content areas assessed, scoring procedures, and technical characteristics.

- **Section Three: Context of Implementation.** This section includes a summary of the policies and procedures followed in implementing the performance assessment, the resources and help available to the education agency or school personnel for developing and implementing the assessment, and the coordination (or lack thereof) between performance assessments and other tests, reforms, and organizational changes.

- **Section Four: The Use and Impact of Performance Assessment.** This section describes the uses of the performance assessments by teachers and students at the sample school. In addition, it documents the school community's evaluation of the usefulness and quality of the assessments and the impact of the assessments on the teaching and learning processes at the school.

Cross-Case Analysis

The second phase of our data analysis focused on extracting and reorganizing information from our case study write-ups into a cross-case comparative format. Based upon the case-study data and the theoretical, empirical, and policy papers we collected, we developed a categorization system for each of our major variables — performance assessments, facilitators and barriers, and teaching and learning processes and outcomes. Next, we organized the data from each case-study into the categorization system. After the categorization exercise, we identified both common patterns and unique features in our data in order to:

- Develop a taxonomy of performance assessments;

- Illustrate and discuss the facilitators and barriers in assessment reform at different levels of educational organization;

- Identify and discuss the concept of "teacher appropriation" of performance assessments as a prerequisite to meaningful changes in teaching practice; and

---

[2] The full case studies appear in Volume II, Studies of Education Reform: Assessment of Student Performance - Case Studies. Only summary case studies are presented in this volume.

- Catalog and describe the impact of different performance assessments on teaching and learning at sample schools.

Thus, our cross-case analysis report comprises four sections.

In the first section we organized, when appropriate, the data according to the organizational level at which the performance assessment system was initiated, developed, or implemented (national, state, district, or school). We developed this organizational scheme to enable us to identify and understand the systematic differences among performance assessments developed and implemented by different levels of education authority.

Similarly, in the second section, we organized the data by the level of initiation, development, or implementation. We isolated and analyzed the organizational factors that influenced the development and implementation of performance assessments at different levels of education authority. In the third section, we organized the data in terms of the facilitators and barriers that affect teachers' ability and willingness to work with the subject performance assessment. Teachers' "appropriation" of the assessment is seen as a necessary prerequisite of meaningful changes in teaching and, hence, learning. These two sections together are keyed to the second study objective.

In the last section (keyed to the third objective), we were interested in the impact of the performance assessment at the local school level. Thus, we organized the data according to different categories of performance assessments, as we wanted to investigate the effects of different types of performance assessments on the teaching and learning processes and outcomes at the local school.

Our approach to data analysis was primarily inductive, and our findings are offered as informed hypotheses that merit further investigation.

Data Interpretation

A research design such as the one we used has strengths, but it also necessarily imposes certain limitations on the interpretations that can be drawn from the data. We briefly discuss five general limitations of our study. (Specific limitations to our analyses are discussed in the appropriate chapters in this report.)

First, our taxonomy of performance assessments is based upon a limited sample of performance assessments. Although we attempted to obtain a representative sample of performance assessments, we are not certain that the assessments initiated at the district and school levels are, in fact, representative of all district- and school-initiated `erformance assessments. Hence, our taxonomic scheme may not be accurate and must be viewed as a work in progress.

Second, a comprehensive description and analysis of each of the performance assessments in our sample was not possible. It was beyond the scope of this study to collect the massive amounts of data required for conducting such an analysis.

Third, our findings regarding the facilitators and barriers in assessment reform, especially at the national-, and state-levels may be less comprehensive than for those at the district- and school-levels. This limitation stems from the local-level emphasis of our study. We collected information regarding national- and state-level assessment reform from documents and general, as opposed to detailed and probing, interviews. In addition, we did not conduct in-person interviews with state officials and researchers involved in national-level efforts as we did with district- and school-level personnel.

Fourth, our findings regarding the impact of national-, state-, and district-initiated performance assessments are valid only for the schools included in this study; the results obtained for a particular school cannot be generalized to other schools involved with the same performance assessments.

Finally, interviewees' opinions regarding impact of and problems with performance assessments signal the existence of those impacts and problems, but the absence of such opinions does not necessarily suggest the absence of impact of or problems with performance assessments.

ATTACHMENT A

Research Questions

The major research questions include the following:

Key Characteristics of Performance Assessments

- What are the key characteristics of these new assessment strategies? How do these new approaches differ from traditional practices and from prior practice in particular sites?

- What key characteristics cut across successful approaches? What characteristics are missing from less successful programs in this area? Why are particular aspects of the model approach especially important?

- What are the purposes of these new assessment approaches? Are those aims different from traditional practice and from prior practice in particular sites?

- What role was played by research, research-based knowledge, and other information in the design of these new assessment approaches? What evidence documents that role?

- What are the content and skills coverages of the assessment? What are the cognitive demands of the assessments?

- What methods are available to measure students' sustained performance? What types of material (and how much) should be kept in portfolios at what age or grade levels? What types of assessment measures can be developed from them? For example, what types of material are needed to reflect: (1) depth and breadth of understanding in various subject matter areas; (2) ability to use resources; (3) ability to work productively with peers; (4) individual productivity; and (5) ability to respond appropriately to constructive review and criticism? What are the best arrangements for keeping portfolios or for deciding what they are to contain or when and if they should be cleaned out?

Facilitators and Barriers in Assessment Reform

- What are the circumstances that permitted or encouraged the development and implementation of new assessment strategies? To what degree, and how, can these or similar conditions be reproduced in other settings? How must different approaches be adapted to particular settings?

- What were the principal incentives for reform? What have been the major barriers to the initiation, development, implementation, and sustenance of the reforms, and

how have those been overcome? What Federal, state, district, or school policies or practices facilitate or inhibit these reforms?

- How have curriculum and instruction been affected by performance assessments? Did a drive for curricular and instructional reform give impetus to the development and adoption of performance assessments?

- Who should be involved in the development of assessments such as performance assessments or portfolios? Where is this developmental work best undertaken? School site? District? State? For example, what types of cooperation and support from the State level are most important for a district interested in developing a new assessment of student performance?

- What strategies are required to get schools organized logistically to develop and implement new forms of student assessment in the schools? What are successful strategies for involving and informing students and parents and other members of the community?

- What types of training are needed for the various individuals involved in these new assessments? What are the components of training and what training approaches are most effective? What are the advantages and disadvantages of alternative types and sources of this training? What training and resource support are needed for change in instructional strategies and curriculum?

- What technical factors need to be considered? How have these been dealt with most successfully?

- What information sources are available for those interested in either developing their own new forms of assessment or adapting existing assessments that have been developed by others as part of their school or curriculum reform? What technical support is available? What was most helpful and what seemed to be missing?

- What resources were required to design, develop, implement or sustain these reforms, including staff time, staff training and support, space, materials and supplies? If extra funds were required, how much extra was needed, what was the source of those funds, and how were they obtained? How are total costs and extra costs related to the number of pupils served?

- What role was played by research, research-based knowledge, and other information in the implementation of these reforms? What evidence documents that role?

Impact of Performance Assessments on Teaching and Learning

- What strategies and approaches have been developed to assess the impact of these reforms? How do these approaches separate the impact of the reforms from the impact of other factors that might affect outcomes? How can these assessments be used to refine reforms?

- What has been the impact of these new assessments, particularly the impact on students, and especially the impact on student performance?

- What impact does the form of a test have on what students learn? What impact does the form of a test have on the quality, quantity, timeliness, and continuity of feedback to students, to their teachers, and to their parents? What information do students, teachers, and parents receive from these assessments about how students can improve their performance?

- What impact does the form of a test have on curriculum and instruction? Do tests that emphasize the mere recall of facts lead to a curriculum that emphasizes routine memorization? Do tests that require students to develop their own solutions and strategies lead to a curriculum emphasizing problem solving and higher-order thinking skills?

- What are the anticipated and unanticipated benefits and difficulties associated with this reform? How can those benefits be reproduced and those difficulties be avoided in other jurisdictions wishing to implement similar reforms?